KOM 6115 HW2

In the following questions, assume that the mathematical model of an MDP (S, A, p_s, r, γ) is given in the following form.

- State set S is represented by integers $S = \{1, \dots, n\}$,
- Action set A is represented by integers $A = \{1, \dots, m\}$,
- State transition probabilities p_s are given by a 3D array P in the form $p_{ijk} = p_s(s^j | s^i, a^k)$,
- Rewards r are given by a 3D array R in the form $r_{ij\,k} = r(s^i, s^j, a^k)$.

Note that since the sets S and R are composed of consecutive integers starting from 1, it is enough to known m and n to define these sets. m and n can be easily obtained from the dimensions of the matrix P or R.

1) Consider a finite horizon MPD $(S, A, p_s, r, 1)$ with the final time T.
   a) Write a MATLAB function that takes arrays P, R, and the horizon T corresponding to the finite horizon MDP and computes the optimal value function and policy of this MDP. The function should return two values. One is a 2D matrix V composed of the elements v_{it} of the time-dependent optimal value function computed. The other is PI composed of the elements of the time-dependent optimal policy π_{it} computed.
   b) Find the optimal value function and policy of the Stochastic Roller problem studied in the class for T=10 using the function you wrote in part a). Note that you should represent the states L, M, H, and the actions spin and don't spin using integers.

2) Consider an infinite horizon MPD (S, A, p_s, r, γ)
   a) Like the first question, write a MATLAB function that takes the data of the infinite horizon MDP (this time instead of T it should be taking γ) and returns the optimal value function and the policy using Dynamic Programming iteration. But this time the optimal value function and policy should be 1D arrays since they are time-independent.
   b) Find the optimal value function and the policy of the stochastic roller problem for the infinite horizon with $\gamma = 0.85$.
   c) Repeat part a) but this time obtain the solution using the LP formulation. Your function should construct the matrices of the LP problem in the standard form using the formulas derived in the class and solve them using *linprog* solver of MATLAB. The solver comes with the optimization toolbox.
   d) Solve the roller problems studied in part b) using the new function you wrote in part c). Does it give the same result?

3) Consider a finite horizon LQR problem defined by the tuple $(A, B, Q, R, \Sigma$,T).
   a) Write a MATLAB function that takes the data of problem and returns the parameters of the optimal value function (P_t, c_t) and the optimal control gains (K_t) which depends on time. Use cell arrays for return values such that each cell will contain data corresponding to a time t.
   b) For T=30, use your function to find optimal value function parameters and control gains of the inverted pendulum problem whose data was given in the first homework.
   c) Write a MATLAB function that takes controller gains, the problem data and initial states. The function should simulate the system for the controller gains provided and return matrices X and U which contain state and input trajectories, respectively, obtained from the simulation.
   d) Use the function wrote in part c) to simulate the system for the solution found in part b). Plot the state and input trajectories to different figures. Note: you can choose initial conditions arbitrarily as long as they are not close to zero.

4) Consider an infinite horizon LQR problem defined by the tuple (A, B, Q, R, Σ).
   a) Write a MATLAB function that takes the data of the problem and returns the optimal value function parameters (P, c) and the optimal controller gain (K), which are independent of time, using Dynamic Programming iteration.
   b) Use your function to find optimal value function parameters and control gains of the inverted pendulum problem whose data was given in the first homework.
   c) Like Question 3), simulate the system and plot state and input trajectories to different figures.